

Jet substructure in ATLAS

David W. Miller, on behalf of the ATLAS Collaboration
 SLAC National Accelerator Laboratory, Menlo Park, CA USA*

Measurements are presented of the jet invariant mass and substructure in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector using an integrated luminosity of 37 pb^{-1} . These results exercise the tools for distinguishing the signatures of new boosted massive particles in the hadronic final state. Two “fat” jet algorithms are used, along with the filtering jet grooming technique that was pioneered in ATLAS. New jet substructure observables are compared for the first time to data at the LHC. Finally, a sample of candidate boosted top quark events collected in the 2010 data is analyzed in detail for the jet substructure properties of hadronic “top-jets” in the final state. These measurements demonstrate not only our excellent understanding of QCD in a new energy regime but open the path to using complex jet substructure observables in the search for new physics.

I. INTRODUCTION

The hadronic final state may be explored in terms of the structure and shape of the hadronic energy flow via the internal hard and soft structure of jets – jet “substructure.” Such an approach seeks to measure important aspects of QCD and to identify patterns and topological features that are potentially indicative of the presence of new physics.

Hadronic decays of heavy particles such as W bosons [1–3], top quarks [4–8], the Higgs boson [9, 10], and potential new particles [11], may be collimated into a single heavy jet characterized by distinct substructure and large mass. In this analysis, two “fat” jet algorithms, along with the filtering jet grooming technique developed in the search for a boosted Higgs boson, are used to measure the jet mass and substructure. Results are presented at the particle level and compared to three Monte Carlo simulation programs. A first measurement of the jet mass scale uncertainty in ATLAS is made, and consequences for searches based on jet mass are discussed. An additional jet substructure observable, the first k_t splitting scale or $\sqrt{d_{12}}$, is measured for the first time at the LHC. Finally, a sample of candidate boosted top quark events collected in the 2010 data is analyzed in detail for the jet substructure properties of hadronic “top-jets” in the final state.

II. DATA SELECTION AND MONTE CARLO SAMPLES

The ATLAS detector [12] provides nearly full solid angle coverage of the collision point with tracking detectors, calorimeters and muon chambers. Of these multiple subsystems the most relevant to the analyses presented here are the inner detector [13], barrel and endcap calorimeters [14, 15] and trigger [40].

The entire 2010 dataset is used for the analysis of jet substructure, corresponding to an integrated luminosity of approximately $37 \pm 1.3 \text{ pb}^{-1}$. The uncertainty of 3.4% on the luminosity is obtained from the latest luminosity calibration using the full set of van der Meer scans performed in 2010 [16].

A sample of high- p_T jets is selected via a single jet trigger with a nominal transverse energy threshold of 95 GeV at the electromagnetic scale. The offline selection requires that the highest p_T jet be within $|\eta| < 2.0$ and to have $p_T > 300$ GeV. Events are required to contain exactly one primary vertex in order to reduce the impact of multiple simultaneous proton-proton interactions, although a dedicated study demonstrating the impact of such pile-up is also presented. The anti- k_t jet algorithm [17, 18] used for most of the results presented yields regularly shaped jets that are well-suited for calibration. The Cambridge-Aachen (C/A) jet algorithm [19, 20] is also very appropriate for measurements of jet substructure in two-body decays such as that expected from a boosted Higgs decaying to two b quarks. Jet radius parameters of $R = 1.0$ and $R = 1.2$ are used for the anti- k_t and C/A algorithms, respectively. Jet quality criteria such as the timing of calorimeters cells, the electromagnetic energy fraction, and pulse shape information are used to select only those events with well measured jets. These criteria are designed to remove events that are likely to have contamination due to beam-related backgrounds or

*Now at The University Of Chicago, Enrico Fermi Institute.

detector defects. Since the primary vertex multiplicity is different between data and MC due to event pile-up, a re-weighting factor is applied to the MC prior to event selection.

The data are compared to events simulated using two different approaches to MC event generation. The first uses direct perturbative calculation of the cross-section matrix elements in powers of the strong coupling constant, α_s . For QCD jet production, this approach is used by ALPGEN [21] and performed at leading-order (LO) in α_s for each relevant partonic subprocess. The second approach implements a sampling of the phase-space available for gluon emission with some suitable approximations. The simulation programs PYTHIA [22] and HERWIG++ [23] both implement this approach for QCD jet production and use LO perturbative calculations of matrix elements for $2 \rightarrow 2$ processes and rely on the parton shower to produce the equivalent of multi-parton final states. In the analysis presented here, we utilize ALPGEN 2.13 with exact matrix element calculations up to $n \leq 6$ partons and interfaced to HERWIG 6.510 [24] to provide the parton shower and hadronization model and with JIMMY 4.31 [25] for the underlying event model. Comparisons are also made to PYTHIA 6.423 and HERWIG++ 2.4 which provide shower models that are p_T^2 ordered and angular ordered, respectively. Leading-order parton density functions are taken from the MRST2007 LO* [26, 27] set, unless stated otherwise.

The MC generated samples are passed through a full simulation of the ATLAS detector and trigger [28] based on GEANT4 [29]. The Quark Gluon String (QGSP) model [30] is used for the fragmentation of the nucleus, and the Bertini cascade (BERT) model [31] for the description of the interactions of the hadrons in the medium of the nucleus. Alternative GEANT4 physics lists, using a combination of the FRITIOF [32] and Bertini models and QGSP without Bertini are used as part of the studies to understand the uncertainties on the jet energy and mass scale.

III. JET SUBSTRUCTURE

The jet mass and internal splitting scales are measured for a sample of inclusive jets with $p_T^{\text{jet}} > 300$ GeV [33]. The two jet algorithms, anti- k_t and C/A, utilized for these measurements use wide characteristic radii ($R = 1.0$ and $R = 1.2$, respectively) as this improves efficiency in the case of searches for new hadronically decaying heavy boosted objects. Furthermore, the jet filtering procedure introduced in Refs. [9, 10] is applied to the C/A algorithm and the resulting “groomed” jet mass is measured.

Whereas the jet mass is the four-momentum sum of the clustered jet constituents, the k_\perp splitting scale is defined by first reclustering the constituents of the jet with the k_\perp algorithm [34, 35]. The final recombination step defines the splitting scale variable as

$$\sqrt{d_{12}} = \min(p_{T,i}, p_{T,j}) \times \delta R_{i,j}; \quad (1)$$

$$\delta R_{i,j} = \sqrt{d\phi_{i,j}^2 + dy_{i,j}^2}, \quad (2)$$

where i, j represent the two proto-jets combined at the final step of the k_t algorithm. Due to the definition of the distance metric for k_t , the last recombination will often represent the two most widely separated and highest $p_{T,\text{jet}}$ constituents. Consequently, for a two-body heavy particle decay the final clustering step will usually be to combine those two decay products. The parameter $\sqrt{d_{12}}$ can therefore be used to distinguish heavy particle decays, which tend to be reasonably symmetric, from largely asymmetric QCD splittings. The expected value for a heavy particle decay is approximately $m/2$ whereas plain QCD jets will tend to have values $\lesssim 20$ GeV with a long continuous tail to high $\sqrt{d_{12}}$.

For C/A jets, the filtering jet grooming procedure aims to identify splittings in a jet which each contribute significantly to the jet invariant mass. The procedure is described in detail in Ref. [33]. At each step during the C/A clustering the masses of the proto-jets to be combined are recorded. This clustering history is then searched in reverse for a step at which the jet mass falls by approximately one-third while the p_T of the proto-jets is roughly symmetric. The jet is then reclustered with a smaller jet radius (typically $R = 0.3$), and if more than two subjets are found the three highest p_T subjets are retained and used to calculate the final jet kinematics.

The jet mass and splitting scale are corrected for detector effects via a bin-by-bin unfolding procedure. The final bin sizes are chosen such that bin migrations are small and the per-bin purity [41] is greater than 50%.

Fully corrected hadron-level distributions of the jet mass and splitting scale are shown in Figures 1-2. Both the unfiltered C/A and filtered C/A jet mass distributions are shown in Figure 1. HERWIG++ predicts slightly more massive jets than supported by the data whereas PYTHIA and HERWIG/JIMMY yield measurements which bracket the data and agree to within systematic uncertainties in all bins. Very notably, the differences between the mass distributions predicted by the various MC programs are greatly reduced after the filtering procedure

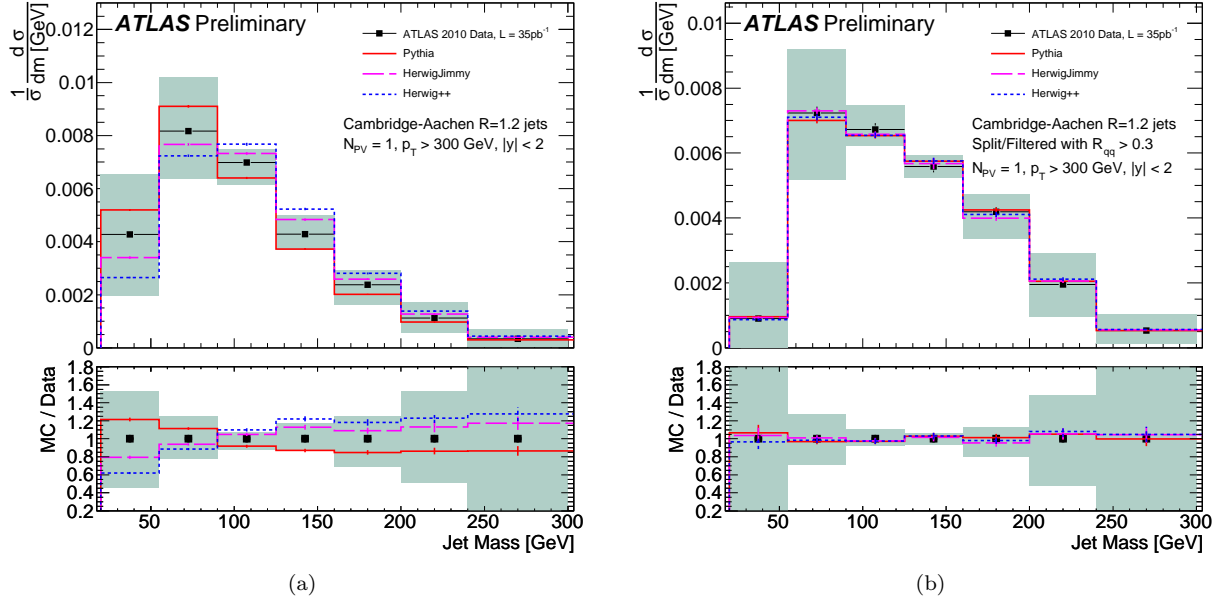


FIG. 1: Invariant mass spectrum of Cambridge-Aachen jets with $p_T > 300$ GeV and $|y| < 2$ (a) before and (b) after the splitting and filtering procedure has been applied. Both distributions are fully corrected for detector effects, systematic uncertainties are depicted by the shaded band.

as shown in Figure 1(a). This observations suggests that the jet mass after the filtering procedure accurately represents the true hard components within the jets which are modeled well by the simulation.

Figure 2 depicts both the fully corrected mass distribution and the k_t splitting scale distribution for anti- k_t , $R = 1.0$ jets. A similar conclusion may be drawn in this case as in the C/A jets, wherein HERWIG++ tends to predict slightly more massive jets on average than observed in the data. However, in this case, the high mass tail of the distribution is actually better reflected in the HERWIG/JIMMY and HERWIG++ MC simulations than in PYTHIA.

Figure 3 exemplifies the potential for adverse pile-up effects and the power of grooming techniques such as jet filtering. The increase in the mean jet invariant mass of calorimeter jets is measured as a function of the number of additional interactions in the event, N_{PV} . Figure 3(a) highlights the dependence of this increase on the radius of the jet algorithm, growing linearly with R in the case of zero pile-up, whereas the slope as a function of pile-up shows an R^3 dependence. This behavior can be qualitatively understood by the growth in jet area as R^2 , and the contribution of additional particles to the jet mass scales with the distance between them $\simeq R/2$ yielding another power of R . Figure 3(b) shows the rise in the mean jet mass as a function of N_{PV} before and after filtering. This procedure restores a flat dependence of the jet mass on N_{PV} to within uncertainties.

IV. BOOSTED HEAVY PARTICLES

A handful of candidate boosted top quark events have been observed in the 2010 data and provide a unique first opportunity to test the techniques described above *in-situ*. This sample has been collected using the high purity lepton+jets channel, which has a signal-to-background (S/B) ratio on the order 2 to 3 [36]. A subsample was then selected to have $m_{t\bar{t}} > 700$ GeV. These events offer a fertile testing ground for the substructure techniques discussed.

Figure 4 shows a single ATLAS event display for one of the selected candidate boosted top quark events. In this display, anti- k_t jets with $R = 0.4$ are indicated in red whereas large radius anti- k_t jets with $R = 1.0$ are shown in green in the same figure. The leading $R = 0.4$ jet in the event corresponds to the leptonic leg of the top event and has $p_T = 199$ GeV. It is also tagged as a b -jet with high probability via both a track impact parameter-based tagger and via a secondary-vertex + impact parameter tagger. The leading $R = 0.4$ jet in the opposite hemisphere, corresponding to the hadronic leg of the event has $p_T = 157$ GeV and no b -tag, whereas the second leading jet in the hemisphere of the hadronic candidate has a strong tag. When re-clustering the

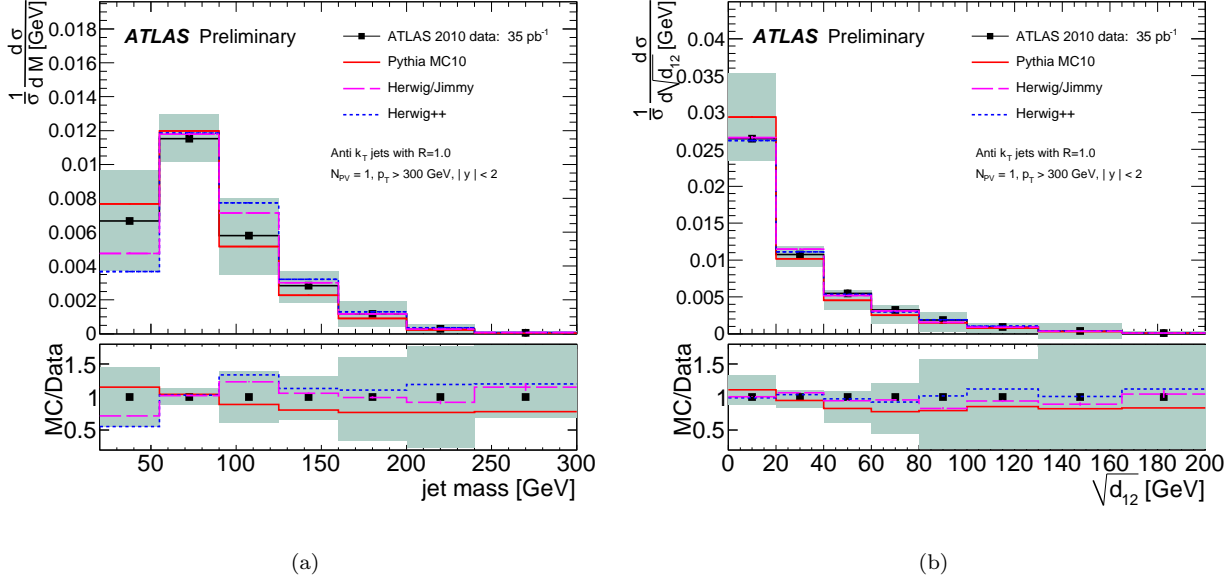


FIG. 2: (a) Invariant mass spectrum of anti- k_t jets with $p_T > 300$ GeV and $|y| < 2$ and (b) $\sqrt{d_{12}}$ distribution for the same jets. Both distributions are fully corrected for detector effects, systematic uncertainties are depicted by the shaded band.

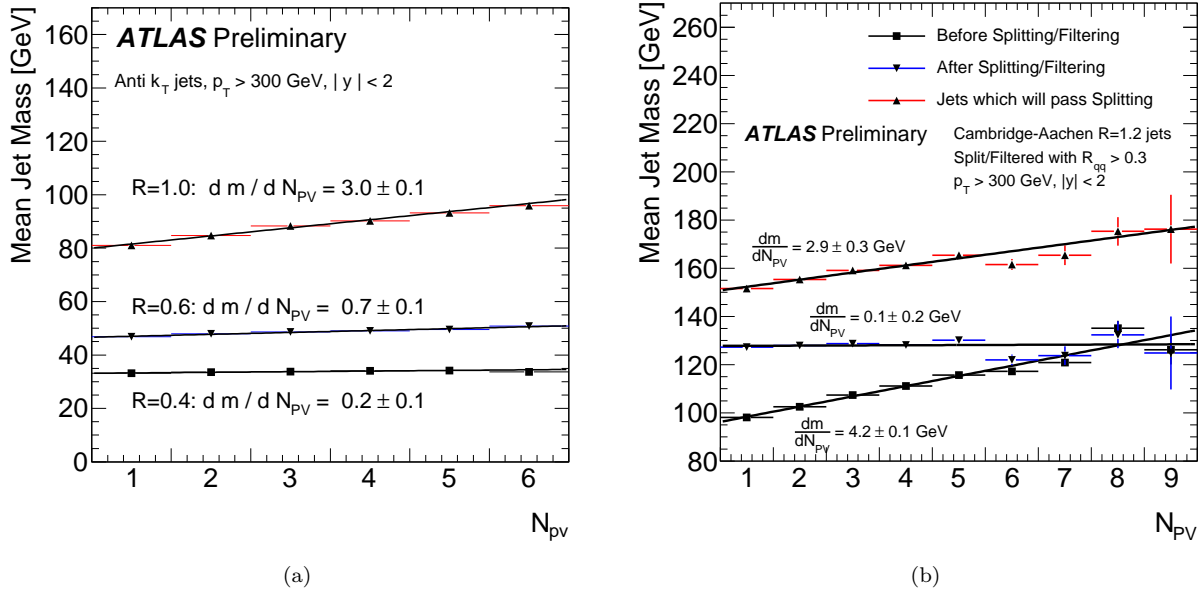


FIG. 3: The mean mass for jets with $p_T > 300$ GeV as a function of the number of primary vertices (N_{PV}) identified in the event. Comparisons show the effect for anti- k_t jets with different resolution parameters (a) and C/A, $R = 1.2$ jets with and without splitting and filtering procedure (b). Each set of points is fitted with a straight line.

event with $R = 1.0$, the jet corresponding to the hadronic leg has $p_T = 327$ GeV and $m^{\text{jet}} = 206$ GeV as well as $\sqrt{d_{12}} = 110$ GeV and $\sqrt{d_{23}} = 40$ GeV. The value of $\sqrt{d_{23}}$ is suggestive of the presence of a boosted hadronically decaying W boson reconstructed within this jet. These observations render this event to be a strong candidate for a boosted hadronically decaying top quark event.

Lastly, a sample of events selected to contain a $W \rightarrow \ell \nu$ candidate consistent with having a W boson $p_T^W > 200$ GeV is analyzed [37]. The jet mass distribution of filtered C/A, $R = 1.2$ jets with $p_T > 180$ GeV and

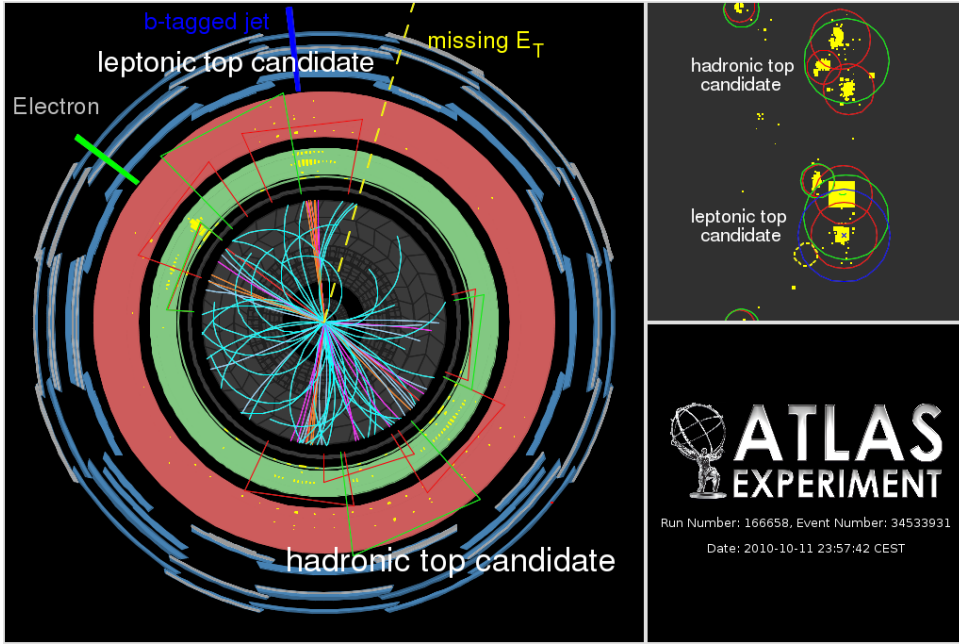


FIG. 4: Leptonic top candidate formed by high p_T electron (145 GeV), moderate E_T^{miss} , and a b -tagged jet. When reclustered with $R = 1.0$ the leptonic candidate acquires a large p_T , mass and $1 \rightarrow 2$ splitting scale as it absorbs the electron. Three jets are identified with the hadronic top quark. When reclustered with $R = 1.0$, these jets merge into a single jet with $m_{jet} = 197$ GeV, $\sqrt{d_{12}} = 110$ GeV, and $\sqrt{d_{23}} = 40$ GeV. Jets indicated in red correspond to $R = 0.4$, jets in green to $R = 1.0$.

$\Delta\phi_{W,jet} > 1.2$ in these events is shown in Figure 5. A clear mass peak near the W mass indicates the presence of boosted hadronic W 's. The three main contributions to these events are $t\bar{t}$ (generated with MC@NLO+HERWIG/JIMMY [38, 39]), W +jets (generated with ALPGEN+HERWIG/JIMMY), and WW (generated with HERWIG/JIMMY), all normalized to the highest order cross-section available (see Ref. [37] for more details). The good agreement between data and the various MC simulations suggests both that the tools described above are well described in a complex physics environment and that the systematics are generally well under control.

V. CONCLUSIONS

The study of the hadronic final state is explored in terms of the substructure of hadronic jets via measurements of the jet mass and k_t splitting scales for anti- k_t jets with $R = 1.0$ and C/A jets with $R = 1.2$. These measurements are corrected for detector effects and a full systematic uncertainty evaluation is performed. In addition, the jet mass is measured for C/A jets after a splitting and filtering procedure. In all observables the PYTHIA and HERWIG samples are in agreement with data to within the systematic uncertainties. The HERWIG++ prediction appears to be slightly disfavoured in the unfiltered C/A mass spectra, producing jets with a higher mass than found in data.

The effect of pile-up on jet mass has also been studied. The expected scaling with number of vertices and jet radius is observed. It has also been shown that the filtering procedure applied to C/A jets reduces the impact of pile-up on jet mass to the extent that it is undetectable in the 2010 dataset.

Finally, the applications of these techniques is demonstrated in the tagging of candidate boosted top quarks and the measurement of fully hadronic W decays.

Overall it is clear that ATLAS is capable of delivering measurements of the variables considered in this study and that these observables are well modeled by leading order Monte Carlo. Searches for boosted Higgs bosons, supersymmetric particles, and top-quark resonances will all benefit from utilizing these techniques.

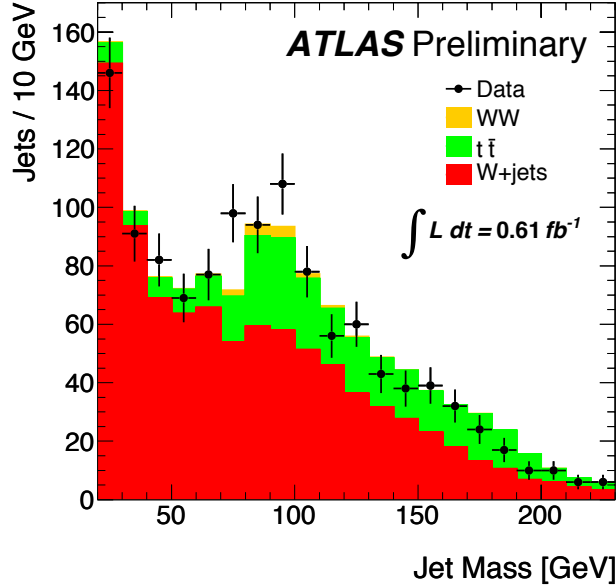


FIG. 5: The jet mass distribution of subjets with $p_T > 180$ GeV in events consistent with a $W \rightarrow l\nu$ boson decay with $p_T > 200$ GeV. The distribution is compared to the uncorrected MC simulation prediction for $t\bar{t}$, W +jets and WW processes.

-
- [1] M. H. Seymour, Z. Phys. **C62**, 127 (1994).
 - [2] J. M. Butterworth, B. E. Cox, and J. R. Forshaw, Phys. Rev. **D65**, 096014 (2002), hep-ph/0201098.
 - [3] ATLAS Collaboration (2009), hep-ex/0901.0512.
 - [4] ATLAS Collaboration, [ATL-PHYS-PUB-2010-008](#) (2010).
 - [5] ATLAS Collaboration, [ATL-PHYS-PUB-2009-081](#) (2009).
 - [6] ATLAS Collaboration, [ATL-PHYS-CONF-2008-008](#) (2009).
 - [7] S. Chekanov and J. Proudfoot, Phys. Rev. **D81**, 114038 (2010), hep-ph/1002.3982.
 - [8] S. Chekanov, C. Levy, J. Proudfoot, and R. Yoshida, Phys. Rev. **D82**, 094029 (2010), hep-ph/1009.2749.
 - [9] J. M. Butterworth et al., Phys. Rev. Lett. **100**, 242001 (2008), hep-ph/0802.2470.
 - [10] ATLAS Collaboration, [ATL-PHYS-PUB-2009-088](#) (2009).
 - [11] J. M. Butterworth et al., Phys. Rev. Lett. **103**, 241803 (2009), hep-ph/0906.0728.
 - [12] ATLAS Collaboration, JINST **3**, S08003 (2008).
 - [13] ATLAS Collaboration, European Physical Journal C **70**, 787 (2010), physics.ins-det/1004.5293.
 - [14] ATLAS Collaboration, European Physical Journal C **70**, 755 (2010), physics.ins-det/1002.4189.
 - [15] ATLAS Collaboration, European Physical Journal C **70**, 1193 (2010), physics.ins-det/1007.5423.
 - [16] ATLAS Collaboration, [ATLAS-CONF-2011-011](#) (2011).
 - [17] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), 0802.1189.
 - [18] M. Cacciari and G. P. Salam, Phys. Lett. B **641**, 57 (2006), hep-ph/0707.1378.
 - [19] Y. L. Dokshitzer et al., JHEP **08**, 001 (1997), hep-ph/9707323.
 - [20] M. Wobisch and T. Wengler (1998), hep-ph/9907280.
 - [21] M. L. Mangano et al., JHEP **07**, 001 (2003), hep-ph/0206293.
 - [22] T. Sjöstrand, Computer Physics Communications **135**, 238 (2001), hep-ph/0010017.
 - [23] M. Bähr et al., European Physical Journal C **58**, 639 (2008), hep-ph/0803.0883.
 - [24] G. Corcella et al., Journal of High Energy Physics **2001**, 010 (2001), hep-ph/0011363.
 - [25] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, Z. Phys. C **72**, 637 (1996), hep-ph/9601371.
 - [26] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. **C63**, 189 (2009), hep-ph/0901.0002.
 - [27] A. Sherstnev and R. S. Thorne, Eur. Phys. J. **C55**, 553 (2008), hep-ph/0711.2473.
 - [28] ATLAS Collaboration, European Physical Journal C **70**, 823 (2010).
 - [29] S. Agostinelli et al. (GEANT4), Nucl. Instrum. Meth. A **506**, 250 (2003).
 - [30] G. Folger and J. P. Wellisch (2003), nucl-th/0306007.
 - [31] H. W. Bertini, Phys. Rev. A **188**, 1711 (1969).
 - [32] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist, Nucl. Phys. B **281**, 289 (1987).

- [33] ATLAS Collaboration, [ATLAS-CONF-2011-073](#) (2011).
- [34] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, Nucl. Phys. **B406**, 187 (1993).
- [35] S. D. Ellis and D. E. Soper, Phys. Rev. **D48**, 3160 (1993), hep-ph/9305266.
- [36] ATLAS Collaboration, European Physical Journal C **71**, 1577 (2011), hep-ex/1012.1792.
- [37] ATLAS Collaboration, [ATLAS-CONF-2011-103](#) (2011).
- [38] S. Frixione and B. R. Webber, JHEP **06**, 029 (2002), hep-ph/0204244.
- [39] S. Frixione, P. Nason, and B. R. Webber, JHEP **08**, 007 (2003), hep-ph/0305252.
- [40] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the center of the detector and the z -axis along the beam pipe. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the beam pipe. Pseudorapidity and rapidity are defined as $\eta = 1/2 \times \ln((|p|+p_z)/(|p|-p_z))$ and $y = 1/2 \times \ln((E+p_z)/(E-p_z))$ respectively.
- [41] The purity here refers to the fraction of reconstructed jets in a bin which were also in the same bin in the truth distribution.